Hierarchical Navigation and Visual Search for Video Keyframe Retrieval

Carles Ventura, Manel Martos, Xavier Giró-i-Nieto, Verónica Vilaplana, and Ferran Marqués

Technical University of Catalonia (UPC), Barcelona, Catalonia / Spain,
{carles.ventura,xavier.giro,veronica.vilaplana,ferran.marques}@upc.edu

Abstract. This work presents a browser that supports two strategies for video browsing: the navigation through visual hierarchies and the retrieval of similar images and objects. The input videos are firstly processed by a keyframe extractor to reduce the temporal redundancy and decrease the number of elements to consider. These generated keyframes are hierarchically clustered with the Hierachical Cellular Tree (HCT) algorithm, an indexing technique that also allows the creation of data structures suitable for browsing. Different clustering criteria are available, in the current implementation, based on four MPEG-7 visual descriptors computed at the global scale. The navigation can directly drive the user to find the video timestamps that best match the query or to a keyframe which is globally or locally similar in visual terms to the query. If this is the case, a visual search engine is also available to find other similar keyframes or regions, also based on MPEG-7 visual descriptors.

Keywords: Video browser, Hierarchical Navigation, Object retrieval, Image retrieval

1 Introduction

As a consequence of recent technology development, large amounts of video data are generated and stored. Accessing these rich portion of data in terms of audiovisual and semantic content is an open research issue with several solution depending on the user needs. This work proposes an hybrid interface that combines both navigation and search tools to provide users with a high degree of flexibility for choosing the most appropriate strategies according to the query nature. Figure 1 shows a screenshot of GOS, the GUI that exploits the presented techniques.

2 Hierarchical Cellular Tree

The Hierarchical Cellular Tree (HCT) algorithm [1] was designed to bring an effective solution for indexing large multimedia databases. The elements are partitioned depending on their relative distances and stored within cells on the

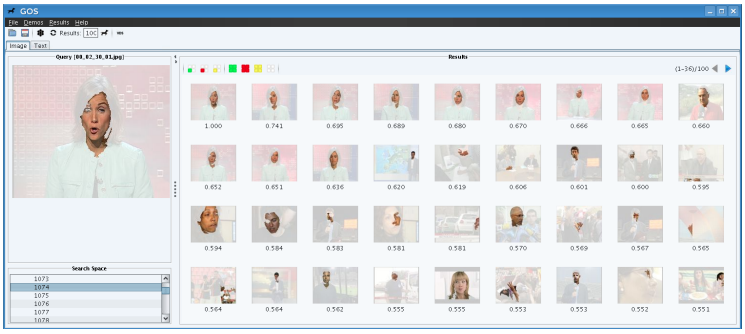

Fig. 1. Screenshot of GOS, the graphical user interface

basis of their similarity. As its name implies, HCT is a hierarchical structure, which consists of one or more levels, and each level in turn holds one or more cells. Each cell has an element as a nucleus, which is contained in a cell in the upper level except for the cell held by the top level. These representative elements are used during the top-down search for item insertions and query requests. Another dynamic cell feature is the cell compactness, which quantifies how focused or compact the clustering for the items within the cell is. This parameter plays a key role in deciding whether or not to perform mitosis (cell splitting) within the cell at any instant.

The hierarchical structure of the HCT allows a multiscale visual navigation since the nucleus of each cell is representative of the underlying elements. Given a level in the HCT, the set of thumbnails built from the nucleus provides the user with visual and intuitive data to decide what path offers greater chances to find the keyframes that better match the query.

3 Visual Search

The proposed system contains a second tool for keyframe search based on visual similarity. Any keyframe can be selected to formulate a query-by-example among the rest of keyframes. This tool helps users to find similar portions of the video as well as co-occurrences of an object through a query by region. In this second case, an interactive segmentation tool [3] is available to navigate through the keyframe regions and define the object of interest.

The same HCT structures used for navigation are also used in visual search at global scale as an index for fast retrieval. This way the same data structure can be exploited manually, though navigation, or by an automatic visual search engine.

Object retrieval requires a segmentation of the keyframes in regions and the extraction of local features from each of these regions. The adopted segmentation algorithm is the Binary Partition Tree [4], which also generates a hierarchy, in this case, of regions. The search algorithm will try to match the query region

with every region and retrieve keyframes according to the best match of one of their regions to the query one.

4 Video browser approach

The problem we aim to solve consists in finding a preselected segment of interest (duration ranging from a few seconds up to 30 seconds) in a one-hour video file within a specified time limit (e.g., within 3 minutes) by interactive search. We have to consider that the preselected segment of interest, i.e. the query clip, is not available for the retrieval system, so the query clip cannot be processed. Therefore, we have to find it by navigating through the video file to which the query clip belongs to.

With this purpose, we have adopted the following strategy, where the first 1-4 step are performed offline to speed up the retrieval process:

1. The test video is previously processed by keyframe extractor in order to work with a lower number of frames which represent the video.
2. Global features are extracted for each keyframe, in particular, the following MPEG-7 visual descriptors [5]: (*i*) Color Structure, (*ii*) Dominant Colors, (*iii*) Color Layout, and (*iv*) Texture Edge Histogram.
3. A HCT is built over each of the extracted visual descriptors, considering as a similarity metric the visual distances recommended in MPEG-7.
4. Local region features are extracted from an automatic segmentation of each keyframe. The considered visual descriptors also come from the MPEG-7 standard: Dominant Colors, Texture Edge Histogram and Contour Shape.
5. Depending on the query clip, the user decides on which visual descriptor the navigation must start. The GUI shows to the user all the elements belonging to level of the hierarchy whose thumbnails fill the screen. Then, the user selects in which subtree is interested. As a consequence, the elements belonging to the level below are shown to the user, who decides whether descending through that subtree or coming back to the upper level.
6. If the navigation drives the user to a keyframe which is globally or locally similar to any of the frames in the segment of interest, the user can launch a visual search process. Results will also be shown through a hierarchy, so a new navigation or search process can start on the new tree.
7. If the navigation drives the user a keyframe of the segment of interest, the timestamp of the keyframe found is used to retrieve their neighbouring keyframes in time.
8. The user finally selects the first and final keyframe of the segment of interest.

References

1. Kiranyaz, S. and Gabbouj, M.: Hierarchical Cellular Tree: An Efficient Indexing Scheme for Content-Based Retrieval on Multimedia Databases. IEEE Transactions on Multimedia, Jan 2007, pp. 102-119.

2. Ventura, C.: Tools for image retrieval in large multimedia databases. Master Thesis at Technical University of Catalonia, September 2011 http://hdl.handle.net/2099.1/13011
3. Giro-i-Nieto, X and Camps, N. and Marques, F.: GAT, a graphical annotation tool for semantic regions. Multimedia Tools and Applications, 46(2):155–174, 2010.
4. Salembier, P. and Garrido, L.: Binary partition tree as an efficient representation for image processing, segmentation, and information retrieval IEEE Transactions on Image Processing, Apr 2000, pp. 561-576.
5. B. S. Manjunath, P. Salembier and T. Sikora: Introduction to MPEG-7, Multimedia Content Description Interface. John Wiley and Sons, Ltd. Jun 2002.